Mathematics for Machine Learning — Probability & Distributions (Supplementary): Gaussian Distribution & Change of Variables/Inverse Transform

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering, Tamkang University

Fall 2023

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

Credits for the resource

- The slides are based on the textbooks:
 - Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.
 - Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra. Wiley. 2019.
- We could partially refer to the monograph: Francesco Orabona: A Modern Introduction to Online Learning. https://arxiv.org/abs/1912.13213

A (1) < A (2) < A (2) </p>

Outline



Gaussian Distribution

- Marginals and Conditionals of Gaussians
- Sums and Linear Transformations

2 Change of Variables

- Distribution Function Technique
- Change of Variables

Fall 2023

Gaussian Distribution

Outline



Gaussian Distribution

- Marginals and Conditionals of Gaussians
- Sums and Linear Transformations

Change of Variables

- Distribution Function Technique
- Change of Variables

< □ > < □ > < □ > < □ > < □ > < □ >

Fall 2023

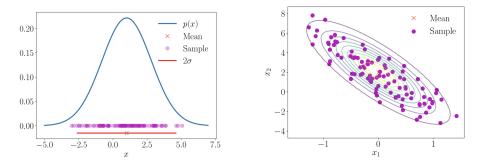
Introduction

- The Gaussian distribution (a.k.s. normal distribution) is the most well-studied probability distribution for continuous-valued random variables.
- Widely used in statistics and machine learning.

・ 同 ト ・ ヨ ト ・ ヨ ト

Gaussian Distribution

Gaussian Distributions Overlaid with Samples



ML Math - Probability & Distributions

Univariate & Multivariate Gaussian

The probability density functions.

Univariate

$$p(x \mid \mu, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\Sigma = \mathbb{V}_X[\mathbf{x}] = \mathsf{Cov}_X[\mathbf{x}, \mathbf{x}].$$

Multivariate

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{D}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right),$$

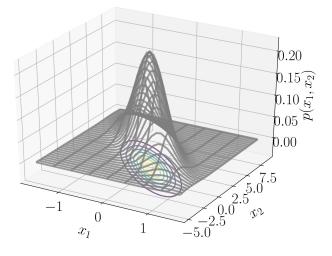
for $\mathbf{x} \in \mathbb{R}^{D}$.

We write $p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ or $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

7/37

Gaussian Distribution

Gaussian distribution of two random variables x_1, x_2 .



э

Gaussian Distribution

Marginals and Conditionals of Gaussians

Outline



Gaussian Distribution

• Marginals and Conditionals of Gaussians

Sums and Linear Transformations

Change of Variables

Joseph C. C. Lin (CSIE, TKU, TW)

- Distribution Function Technique
- Change of Variables

▲ □ ▶ ▲ □ ▶ ▲ □ ▶

Gaussian Distribution

Marginals and Conditionals of Gaussians

Marginals and Conditionals of Gaussians

- Let X, Y be two multivariate random variables.
- Concatenate their states to be $[\mathbf{x}^{\top}, \mathbf{y}^{\top}]$.

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\left[\begin{array}{cc} \boldsymbol{\mu}_{x} \\ \boldsymbol{\mu}_{y} \end{array}\right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{array}\right]\right).$$

where $\Sigma_{xx} = \text{Cov}[\mathbf{x}, \mathbf{x}]$, $\Sigma_{yy} = \text{Cov}[\mathbf{y}, \mathbf{y}]$, $\Sigma_{xy} = \text{Cov}[\mathbf{x}, \mathbf{y}]$.

• By [Bishop 2006], the conditional distribution $p(\mathbf{x} \mid \mathbf{y})$ is also Gaussian.

$$egin{array}{rcl} oldsymbol{
ho}(\mathbf{x} \mid \mathbf{y}) &=& \mathcal{N}(oldsymbol{\mu}_{x\mid y}, oldsymbol{\Sigma}_{x\mid y}) \ oldsymbol{\mu}_{x\mid y} &=& oldsymbol{\mu}_{x} + oldsymbol{\Sigma}_{xy} oldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - oldsymbol{\mu}_{y}) \ oldsymbol{\Sigma}_{x\mid y} &=& oldsymbol{\Sigma}_{xx} - oldsymbol{\Sigma}_{xy} oldsymbol{\Sigma}_{yy}^{-1} oldsymbol{\Sigma}_{yx}. \end{array}$$

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y} = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_{x}, \boldsymbol{\Sigma}_{xx}).$$

r

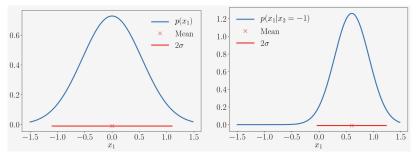
Gaussian Distribution

Marginals and Conditionals of Gaussians

Example

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\left[\begin{array}{cc} 0\\ 2\end{array}\right], \left[\begin{array}{cc} 0.3 & -1\\ -1 & 5\end{array}\right]\right).$$



<ロト < 四ト < 三ト < 三ト

3

Joseph C. C. Lin (CSIE, TKU, TW)

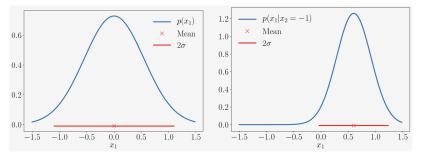
Gaussian Distribution

Marginals and Conditionals of Gaussians

Example

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\left[\begin{array}{cc} 0\\ 2\end{array}\right], \left[\begin{array}{cc} 0.3 & -1\\ -1 & 5\end{array}\right]\right).$$



Conditioned on $x_2 = -1$, $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1-2) = 0.6$

Joseph C. C. Lin (CSIE, TKU, TW)

ML Math - Probability & Distributions

Fall 2023

(日)

11/37

3

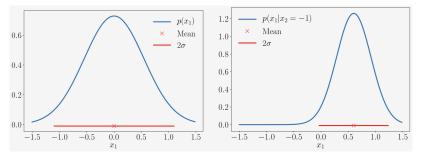
Gaussian Distribution

Marginals and Conditionals of Gaussians

Example

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\left[\begin{array}{cc} 0\\ 2\end{array}\right], \left[\begin{array}{cc} 0.3 & -1\\ -1 & 5\end{array}\right]\right).$$



Conditioned on $x_2 = -1$, $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1-2) = 0.6$ and $\sigma^2_{x_1|x_2=-1} =$

Fall 2023

< 日 > < 同 > < 三 > < 三 > <

11/37

3

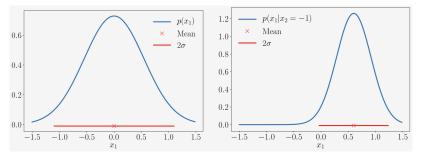
Gaussian Distribution

Marginals and Conditionals of Gaussians

Example

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\left[\begin{array}{cc} 0\\ 2\end{array}\right], \left[\begin{array}{cc} 0.3 & -1\\ -1 & 5\end{array}\right]\right).$$



Conditioned on $x_2 = -1$, $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1-2) = 0.6$ and $\sigma^2_{x_1|x_2=-1} = 0.3 - (-1) \cdot 0.2 \cdot (-1) = 0.1$. Thus, $p(x_1 \mid x_2 = -1) =$

Fall 2023

11/37

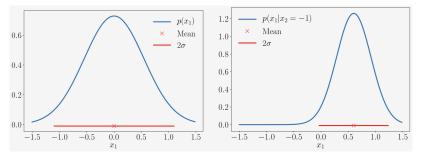
Gaussian Distribution

Marginals and Conditionals of Gaussians

Example

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\left[\begin{array}{cc} 0\\ 2\end{array}\right], \left[\begin{array}{cc} 0.3 & -1\\ -1 & 5\end{array}\right]\right).$$



Conditioned on $x_2 = -1$, $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1-2) = 0.6$ and $\sigma^2_{x_1|x_2=-1} = 0.3 - (-1) \cdot 0.2 \cdot (-1) = 0.1$. Thus, $p(x_1 \mid x_2 = -1) = \mathcal{N}(0.6, 0.1)$,

Fall 2023

11/37

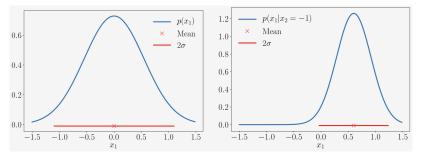
Gaussian Distribution

Marginals and Conditionals of Gaussians

Example

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\left[\begin{array}{cc} 0\\ 2\end{array}\right], \left[\begin{array}{cc} 0.3 & -1\\ -1 & 5\end{array}\right]\right).$$



Conditioned on $x_2 = -1$, $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1-2) = 0.6$ and $\sigma^2_{x_1|x_2=-1} = 0.3 - (-1) \cdot 0.2 \cdot (-1) = 0.1$. Thus, $p(x_1 \mid x_2 = -1) = \mathcal{N}(0.6, 0.1), \quad p(x_1) =$

Fall 2023

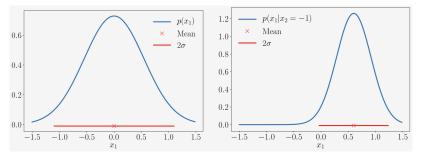
Gaussian Distribution

Marginals and Conditionals of Gaussians

Example

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\left[\begin{array}{cc} 0\\ 2\end{array}\right], \left[\begin{array}{cc} 0.3 & -1\\ -1 & 5\end{array}\right]\right).$$



Conditioned on $x_2 = -1$, $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1-2) = 0.6$ and $\sigma^2_{x_1|x_2=-1} = 0.3 - (-1) \cdot 0.2 \cdot (-1) = 0.1$. Thus, $p(x_1 \mid x_2 = -1) = \mathcal{N}(0.6, 0.1)$, $p(x_1) = \mathcal{N}(0, 0.3)$.

Gaussian Distribution

Sums and Linear Transformations

Outline



Gaussian Distribution

• Marginals and Conditionals of Gaussians

Sums and Linear Transformations

Change of Variables

Joseph C. C. Lin (CSIE, TKU, TW)

- Distribution Function Technique
- Change of Variables

< □ > < □ > < □ > < □ > < □ > < □ >

ML Math - Probability & Distributions Gaussian Distribution

Sums and Linear Transformations

Sum of Gaussians

Say X, Y are two independent Gaussian random variables with

 $X \sim \mathcal{N}(\mu_x, \Sigma_x)$ and $Y \sim \mathcal{N}(\mu_y, \Sigma_y).$

イロト 不得 トイヨト イヨト 二日

ML Math - Probability & Distributions Gaussian Distribution Sums and Linear Transformations

Sum of Gaussians

Say X, Y are two independent Gaussian random variables with

$$X \sim \mathcal{N}(oldsymbol{\mu}_{\mathsf{x}}, oldsymbol{\Sigma}_{\mathsf{x}}) \hspace{0.2cm} ext{and} \hspace{0.2cm} Y \sim \mathcal{N}(oldsymbol{\mu}_{\mathsf{y}}, oldsymbol{\Sigma}_{\mathsf{y}}).$$

• independency:
$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$$
.

э

イロト イボト イヨト イヨト

ML Math - Probability & Distributions Gaussian Distribution Sums and Linear Transformations

Sum of Gaussians

Say X, Y are two independent Gaussian random variables with

$$X \sim \mathcal{N}(oldsymbol{\mu}_x, oldsymbol{\Sigma}_x) \hspace{0.2cm} ext{and} \hspace{0.2cm} Y \sim \mathcal{N}(oldsymbol{\mu}_y, oldsymbol{\Sigma}_y).$$

• independency:
$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$$
.

Then X + Y is also a Gaussian distribution with

$$X + Y \sim \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)$$

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

(日)

э

ML Math - Probability & Distributions Gaussian Distribution Sums and Linear Transformations

Sum of Gaussians

Say X, Y are two independent Gaussian random variables with

$$X \sim \mathcal{N}(oldsymbol{\mu}_{\mathsf{x}}, oldsymbol{\Sigma}_{\mathsf{x}}) \hspace{0.2cm} ext{and} \hspace{0.2cm} Y \sim \mathcal{N}(oldsymbol{\mu}_{\mathsf{y}}, oldsymbol{\Sigma}_{\mathsf{y}}).$$

• independency:
$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$$
.

Then X + Y is also a Gaussian distribution with

$$X + Y \sim \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)$$

Please recall $\mathbb{E}[\mathbf{x} + \mathbf{y}]$ and $\mathbb{V}[\mathbf{x} + \mathbf{y}]$.

イロト 不得 トイヨト イヨト

3

Joseph C. C. Lin (CSIE, TKU, TW)

Gaussian Distribution

Sums and Linear Transformations

Example

Linear Combination of Gaussians

$$p(a\mathbf{x} + b\mathbf{y}) =$$

イロト イヨト イヨト イヨト

Gaussian Distribution

Sums and Linear Transformations

Example

Linear Combination of Gaussians

$$p(a\mathbf{x} + b\mathbf{y}) = \mathcal{N}(a\mu_x + b\mu_y),$$

э

イロト イヨト イヨト イヨト

Joseph C. C. Lin (CSIE, TKU, TW)

Gaussian Distribution

Sums and Linear Transformations

Example

Linear Combination of Gaussians

$$p(a\mathbf{x} + b\mathbf{y}) = \mathcal{N}(a\mu_x + b\mu_y, \ a^2\Sigma_x + b^2\Sigma_y).$$

イロト イヨト イヨト イヨト

э

Gaussian Distribution

Sums and Linear Transformations

Example

Linear Combination of Gaussians

$$p(a\mathbf{x} + b\mathbf{y}) = \mathcal{N}(a\mu_x + b\mu_y, \ a^2\Sigma_x + b^2\Sigma_y).$$

Theorem [Mixture of Two Univariate Gaussian Densities]

Consider a mixture of two univariate Gaussian densities

$$p(x) = \alpha p_1(x) + (1 - \alpha)p_2(x)$$

for the mixture weight $0 < \alpha < 1$ and $(\mu_1, \sigma_1^2) \neq (\mu_2, \sigma_2^2)$. Then,

$$\mathbb{E}[x] = \alpha \mu_1 + (1 - \alpha) \mu_2 \mathbb{V}[x] = [\alpha \sigma_1^2 + (1 - \alpha) \sigma_2^2] + ([\alpha \mu_1^2 + (1 - \alpha) \mu_2^2] - [\alpha \mu_1 + (1 - \alpha) \mu_2]^2).$$

Joseph C. C. Lin (CSIE, TKU, TW)

Fall 2023

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

14 / 37

э

Gaussian Distribution

Sums and Linear Transformations

Proof of the Theorem

Sketch:

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} xp(x) dx = \int_{-\infty}^{\infty} (\alpha xp_1(x) + (1-\alpha)xp_2(x)) dx$$
$$= \alpha \mu_1 + (1-\alpha)\mu_2.$$

2 $\mathbb{E}[x^2] =$

3

<ロト <回ト < 回ト < 回ト -

Gaussian Distribution

Sums and Linear Transformations

Proof of the Theorem

Sketch:

•
$$\mathbb{E}[x] = \int_{-\infty}^{\infty} xp(x) dx = \int_{-\infty}^{\infty} (\alpha x p_1(x) + (1 - \alpha) x p_2(x)) dx$$

 $= \alpha \mu_1 + (1 - \alpha) \mu_2.$
• $\mathbb{E}[x^2] = \int_{-\infty}^{\infty} x^2 p(x) dx = \int_{-\infty}^{\infty} (\alpha x^2 p_1(x) + (1 - \alpha) x^2 p_2(x)) dx$
 $= \alpha (\mu_1^2 + \sigma_1^2) + (1 - \alpha) (\mu_2^2 + \sigma_2^2).$

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

2

Gaussian Distribution

Sums and Linear Transformations

Proof of the Theorem

Sketch:

E[x] =
$$\int_{-\infty}^{\infty} xp(x) dx = \int_{-\infty}^{\infty} (\alpha x p_1(x) + (1 - \alpha) x p_2(x)) dx$$

 = $\alpha \mu_1 + (1 - \alpha) \mu_2$.

 E[x²] = $\int_{-\infty}^{\infty} x^2 p(x) dx = \int_{-\infty}^{\infty} (\alpha x^2 p_1(x) + (1 - \alpha) x^2 p_2(x)) dx$
 = $\alpha (\mu_1^2 + \sigma_1^2) + (1 - \alpha) (\mu_2^2 + \sigma_2^2)$.
 • Recall: $\mathbb{V}_X[x] = \mathbb{E}_X[x^2] - (\mathbb{E}_X[x])^2$.

2

Gaussian Distribution

Sums and Linear Transformations

Proof of the Theorem

Sketch:

a
$$\mathbb{E}[x] = \int_{-\infty}^{\infty} xp(x) dx = \int_{-\infty}^{\infty} (\alpha x p_1(x) + (1 - \alpha) x p_2(x)) dx$$

 $= \alpha \mu_1 + (1 - \alpha) \mu_2.$
a $\mathbb{E}[x^2] = \int_{-\infty}^{\infty} x^2 p(x) dx = \int_{-\infty}^{\infty} (\alpha x^2 p_1(x) + (1 - \alpha) x^2 p_2(x)) dx$
 $= \alpha (\mu_1^2 + \sigma_1^2) + (1 - \alpha) (\mu_2^2 + \sigma_2^2).$
a Recall: $\mathbb{V}_X[x] = \mathbb{E}_X[x^2] - (\mathbb{E}_X[x])^2.$

Using 1 & 2 we can prove the theorem.

イロト イボト イヨト イヨト

э

Gaussian Distribution

Sums and Linear Transformations

Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(oldsymbol{\mu}, oldsymbol{\Sigma})$ and $oldsymbol{y} = oldsymbol{A} oldsymbol{x}$

Joseph C. C. Lin (CSIE, TKU, TW)

• The expectation:
$$\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] =$$

э

(日)

Gaussian Distribution

Sums and Linear Transformations

Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(oldsymbol{\mu}, oldsymbol{\Sigma})$ and $oldsymbol{y} = oldsymbol{A} oldsymbol{x}$

• The expectation: $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\mu$.

(日)

Fall 2023

Joseph C. C. Lin (CSIE, TKU, TW)

Gaussian Distribution

Sums and Linear Transformations

Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(oldsymbol{\mu}, oldsymbol{\Sigma})$ and $oldsymbol{y} = oldsymbol{A}oldsymbol{x}$

- The expectation: $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\mu$.
- The variance: $\mathbb{V}[\textbf{y}] = \mathbb{V}[\textbf{A}\textbf{x}] =$

イロト 不得 トイヨト イヨト

Fall 2023

Joseph C. C. Lin (CSIE, TKU, TW)

Gaussian Distribution

Sums and Linear Transformations

Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(oldsymbol{\mu}, oldsymbol{\Sigma})$ and $oldsymbol{y} = oldsymbol{A}oldsymbol{x}$

Joseph C. C. Lin (CSIE, TKU, TW)

- The expectation: $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\mu$.
- The variance: $\mathbb{V}[\mathbf{y}] = \mathbb{V}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{V}[\mathbf{x}]\mathbf{A}^{\top} = \mathbf{A}\Sigma\mathbf{A}^{\top}$.

イロト 不得 トイヨト イヨト 二日

Fall 2023

Gaussian Distribution

Sums and Linear Transformations

Linear Transformation by a Matrix (1/2)

$(X \sim \mathcal{N}(oldsymbol{\mu}, oldsymbol{\Sigma})$ and $oldsymbol{y} = oldsymbol{A}oldsymbol{x}$

- The expectation: $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\mu$.
- The variance: $\mathbb{V}[\mathbf{y}] = \mathbb{V}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{V}[\mathbf{x}]\mathbf{A}^{\top} = \mathbf{A}\Sigma\mathbf{A}^{\top}$.
- Thus, we have

Joseph C. C. Lin (CSIE, TKU, TW)

イロト 不得下 イヨト イヨト 二日

Fall 2023

Gaussian Distribution

Sums and Linear Transformations

Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(oldsymbol{\mu}, oldsymbol{\Sigma})$ and $oldsymbol{y} = oldsymbol{A}oldsymbol{x}$

- The expectation: $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\mu$.
- The variance: $\mathbb{V}[\mathbf{y}] = \mathbb{V}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{V}[\mathbf{x}]\mathbf{A}^{\top} = \mathbf{A}\Sigma\mathbf{A}^{\top}$.

Thus, we have

 $Y \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top}).$

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.



- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \mathbf{\Sigma}).$
 - Note: A might not be invertible.

< ロ > < 同 > < 回 > < 回 > < 回 > <

Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.



- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \mathbf{\Sigma}).$
 - Note: A might not be invertible.

Joseph C. C. Lin (CSIE, TKU, TW)

ML Math - Probability & Distributions

Fall 2023

イロト 不得 ト イヨト イヨト

3

Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.



• $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \mathbf{\Sigma}).$

Joseph C. C. Lin (CSIE, TKU, TW)

• Note: A might not be invertible.

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^{\top}\mathbf{y} = \mathbf{A}^{\top}\mathbf{A}\mathbf{x}$$

э

イロト イヨト イヨト イヨト

ML Math - Probability & Distributions Gaussian Distribution Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.



• $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \mathbf{\Sigma}).$

Joseph C. C. Lin (CSIE, TKU, TW)

• Note: A might not be invertible.

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^{\top}\mathbf{y} = \mathbf{A}^{\top}\mathbf{A}\mathbf{x} \iff (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y} = \mathbf{x}.$$

э

イロト イヨト イヨト イヨト

Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.



- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \mathbf{\Sigma}).$
 - Note: A might not be invertible.

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^{\top}\mathbf{y} = \mathbf{A}^{\top}\mathbf{A}\mathbf{x} \iff (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y} = \mathbf{x}.$$

• This works even for non-invertible **A**!.

イロト イヨト イヨト イヨト

Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.

 $Y \sim \mathcal{N}(\mu_y, \Sigma)$, $\mathbf{y} = \mathbf{A}\mathbf{x}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$, a full rank $\mathbf{A} \in \mathbb{R}^{M imes N}$, $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \mathbf{\Sigma}).$
 - Note: A might not be invertible.

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^{\top}\mathbf{y} = \mathbf{A}^{\top}\mathbf{A}\mathbf{x} \iff (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y} = \mathbf{x}.$$

• This works even for non-invertible **A**!.

• The variance: $\mathbb{V}[\mathbf{x}] = \mathbb{V}[(\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y}] =$

Joseph C. C. Lin (CSIE, TKU, TW) ML Ma

ML Math - Probability & Distributions

イロン イヨン イヨン

3

Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.

 $Y \sim \mathcal{N}(\mu_y, \Sigma)$, $\mathbf{y} = \mathbf{A}\mathbf{x}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$, a full rank $\mathbf{A} \in \mathbb{R}^{M imes N}$, $M \geq N$

• $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \mathbf{\Sigma}).$

Joseph C. C. Lin (CSIE, TKU, TW)

• Note: A might not be invertible.

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^{\top}\mathbf{y} = \mathbf{A}^{\top}\mathbf{A}\mathbf{x} \iff (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y} = \mathbf{x}.$$

• This works even for non-invertible **A**!.

• The variance: $\mathbb{V}[\mathbf{x}] = \mathbb{V}[(\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y}] = (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\Sigma\mathbf{A}(\mathbf{A}^{\top}\mathbf{A})^{-1}$.

3

・ロト ・ 四ト ・ ヨト ・ ヨト …

Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.

 $Y \sim \mathcal{N}(\mu_y, \Sigma)$, $\mathbf{y} = \mathbf{A}\mathbf{x}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$, a full rank $\mathbf{A} \in \mathbb{R}^{M imes N}$, $M \geq N$

• $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \mathbf{\Sigma}).$

• Note: A might not be invertible.

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^{\top}\mathbf{y} = \mathbf{A}^{\top}\mathbf{A}\mathbf{x} \iff (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y} = \mathbf{x}.$$

• This works even for non-invertible **A**!.

• The variance: $\mathbb{V}[\mathbf{x}] = \mathbb{V}[(\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y}] = (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\Sigma\mathbf{A}(\mathbf{A}^{\top}\mathbf{A})^{-1}.$

• Thus, we have

Joseph C. C. Lin (CSIE, TKU, TW)

イロト 不得下 イヨト イヨト 二日

Sums and Linear Transformations

Linear Transformation by a Matrix (2/2)

Let's consider the reverse transformation.

 $Y \sim \mathcal{N}(\mu_v, \Sigma), \ \mathbf{y} = \mathbf{A} \mathbf{x}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$, a full rank $\mathbf{A} \in \mathbb{R}^{M \times N}, \ M \geq N$

• $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \boldsymbol{\Sigma}).$

• Note: A might not be invertible.

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^{\top}\mathbf{y} = \mathbf{A}^{\top}\mathbf{A}\mathbf{x} \iff (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y} = \mathbf{x}.$$

• This works even for non-invertible **A**!.

• The variance: $\mathbb{V}[\mathbf{x}] = \mathbb{V}[(\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{y}] = (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\Sigma\mathbf{A}(\mathbf{A}^{\top}\mathbf{A})^{-1}$.

Thus, we have

$$X \sim \mathcal{N}((\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\boldsymbol{\mu}_{y}, (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\Sigma\mathbf{A}(\mathbf{A}^{\top}\mathbf{A})^{-1}).$$

Gaussian Distribution

Sums and Linear Transformations

Exercise

Another example of reverse transformation.



- Compute $\mathbb{E}[\mathbf{x}]$.
- Compute $\mathbb{V}[\mathbf{x}]$.
- Derive $X \sim \mathcal{N}(?, ?)$.

イロト 不得 トイヨト イヨト 二日

Gaussian Distribution

Sums and Linear Transformations

A Sampling Approach

We want to obtain samples from a multivariate $\mathcal{N}(\mu, \Sigma)$.

• However, we only have a sampler of $\mathcal{N}(\mathbf{0}, \mathbf{I})$ at hand.

(日)

Gaussian Distribution

Sums and Linear Transformations

A Sampling Approach

We want to obtain samples from a multivariate $\mathcal{N}(\mu, \Sigma)$.

• However, we only have a sampler of $\mathcal{N}(\mathbf{0}, \mathbf{I})$ at hand.

- Assume that we have $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Then, define $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\mu}$, where $\mathbf{A}\mathbf{I}\mathbf{A}^{\top} = \mathbf{A}\mathbf{A}^{\top} = \boldsymbol{\Sigma}$.

・ロト ・四ト ・ヨト ・ヨト

Gaussian Distribution

Sums and Linear Transformations

A Sampling Approach

We want to obtain samples from a multivariate $\mathcal{N}(\mu, \Sigma)$.

• However, we only have a sampler of $\mathcal{N}(\mathbf{0}, \mathbf{I})$ at hand.

- Assume that we have $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Then, define $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\mu}$, where $\mathbf{A}\mathbf{I}\mathbf{A}^{\top} = \mathbf{A}\mathbf{A}^{\top} = \boldsymbol{\Sigma}$.
- To derive **A**:

Joseph C. C. Lin (CSIE, TKU, TW)

Gaussian Distribution

Sums and Linear Transformations

A Sampling Approach

We want to obtain samples from a multivariate $\mathcal{N}(\mu, \Sigma)$.

However, we only have a sampler of N(0, I) at hand.

- Assume that we have $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Then, define $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\mu}$, where $\mathbf{A}\mathbf{I}\mathbf{A}^{\top} = \mathbf{A}\mathbf{A}^{\top} = \boldsymbol{\Sigma}$.
- To derive **A**: Use Cholesky decomposition of the covariance matrix Σ .
 - **A** will be triangular and efficient for computation.

・ロト ・四ト ・ヨト ・ ヨト

Change of Variables

Outline

Gaussian Distribution

- Marginals and Conditionals of Gaussians
- Sums and Linear Transformations

Change of Variables

Joseph C. C. Lin (CSIE, TKU, TW)

2

- Distribution Function Technique
- Change of Variables

< □ > < □ > < □ > < □ > < □ > < □ >

Motivation

Consider the following examples.

- Assuming that X is a random variable distributed according to some well-known distribution, then what is the distribution of X^2 ?
- Assuming that X_1, X_2 are two univariate standard normal distributions, then what is the distribution of $\frac{1}{2}(X_1 + X_2)$?

< 回 > < 三 > < 三 > <

Motivation

Consider the following examples.

- Assuming that X is a random variable distributed according to some well-known distribution, then what is the distribution of X^2 ?
- Assuming that X_1, X_2 are two univariate standard normal distributions, then what is the distribution of $\frac{1}{2}(X_1 + X_2)$?
- What if the transformation is nonlinear?

Motivation

Consider the following examples.

- Assuming that X is a random variable distributed according to some well-known distribution, then what is the distribution of X^2 ?
- Assuming that X_1, X_2 are two univariate standard normal distributions, then what is the distribution of $\frac{1}{2}(X_1 + X_2)$?
- What if the transformation is nonlinear?
 - Closed-form expressions are not readily available.

A (1) < A (2) < A (2) </p>

Change of Variables

Straightforward for Discrete Random Variables

Example: Univariate Random Variables

Given

- A discrete random variable X with pmf Pr[X = x].
- An invertible function U(x).

Consider the transformed random variable Y := U(X) with pmf Pr[Y = y]. Then

Pr[Y = y] = Pr[U(X) = y](transformation of interest) = Pr[X = U⁻¹(y)] (inverse)

where we can observe $x = U^{-1}(y)$.

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの

Change of Variables

Two Approaches

- So far we considered the discrete case (e.g., Pr[X = x]).
- For continuous distributions, we will consider the two approaches:
 Cumulative distribution (Distribution Function Technique).
 - Ohange-of-variable.

イロト 不得 ト イヨト イヨト

Change of Variables

Distribution Function Technique

Outline

Gaussian Distribution

- Marginals and Conditionals of Gaussians
- Sums and Linear Transformations

Change of Variables

Joseph C. C. Lin (CSIE, TKU, TW)

2

- Distribution Function Technique
- Change of Variables

< □ > < □ > < □ > < □ > < □ > < □ >

Change of Variables

Distribution Function Technique

Distribution Function Technique

Note: a cdf of X: $F_X(x) = \Pr[X \le x]$.

Goal: Find the cdf of the random variable Y := U(X)

Ind the cdf

 $F_{Y}(y) = \Pr[Y \leq y].$

2 Differentiating $F_Y(y)$ to get the pdf $f_Y(y)$:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_Y(y).$$

Joseph C. C. Lin (CSIE, TKU, TW)

ML Math - Probability & Distributions

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの Fall 2023

25 / 37

Change of Variables

Distribution Function Technique

Distribution Function Technique

Note: a cdf of X: $F_X(x) = \Pr[X \le x]$.

Goal: Find the cdf of the random variable Y := U(X)

Ind the cdf

 $F_Y(y) = \Pr[Y \leq y].$

2 Differentiating $F_Y(y)$ to get the pdf $f_Y(y)$:

$$f_{\mathbf{Y}}(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_{\mathbf{Y}}(y).$$

Note: The domain of the random variable may have changed!

Joseph C. C. Lin (CSIE, TKU, TW)

ML Math - Probability & Distributions

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの Fall 2023

25 / 37

Change of Variables

Distribution Function Technique

Example

Example

Let X be a continuous random variable with pdf $f_X : [0,1] \mapsto [0,1]$:

$$f_X(x)=3x^2.$$

Goal: Find the pdf of $Y = X^2$.

$F_Y(y) = \Pr[Y \leq y]$

イロト 不得 トイヨト イヨト

Fall 2023

Joseph C. C. Lin (CSIE, TKU, TW)

Change of Variables

Distribution Function Technique

Example

Example

Let X be a continuous random variable with pdf $f_X : [0,1] \mapsto [0,1]$:

$$f_X(x)=3x^2.$$

Goal: Find the pdf of $Y = X^2$.

$$F_Y(y) = \Pr[Y \le y] = \Pr[X^2 \le y]$$
$$= \Pr[X \le y^{\frac{1}{2}}]$$
$$= F_X(y^{\frac{1}{2}})$$

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

(日)

э

Change of Variables

Distribution Function Technique

Example

Example

Let X be a continuous random variable with pdf $f_X : [0,1] \mapsto [0,1]$:

$$f_X(x)=3x^2.$$

Goal: Find the pdf of $Y = X^2$.

$$F_{Y}(y) = \Pr[Y \le y] = \Pr[X^{2} \le y]$$

= $\Pr[X \le y^{\frac{1}{2}}]$
= $F_{X}(y^{\frac{1}{2}}) = \int_{0}^{y^{\frac{1}{2}}} 3t^{2} dt$
= $[t^{3}]_{0}^{y^{\frac{1}{2}}} = y^{\frac{3}{2}}, \quad 0 \le y \le 1.$

Joseph C. C. Lin (CSIE, TKU, TW)

< □ > < □ > < □ > < □ > < □ > < □ >
 Fall 2023

26 / 37

э

Change of Variables

Distribution Function Technique

Example

Example

Let X be a continuous random variable with pdf $f_X : [0,1] \mapsto [0,1]$:

$$f_X(x)=3x^2.$$

Goal: Find the pdf of $Y = X^2$.

$$F_{Y}(y) = \Pr[Y \le y] = \Pr[X^{2} \le y]$$

$$= \Pr[X \le y^{\frac{1}{2}}]$$

$$= F_{X}(y^{\frac{1}{2}}) = \int_{0}^{y^{\frac{1}{2}}} 3t^{2} dt$$
for $0 \le y \le 1$.
$$= [t^{3}]_{0}^{y^{\frac{1}{2}}} = y^{\frac{3}{2}}, \quad 0 \le y \le 1.$$

Joseph C. C. Lin (CSIE, TKU, TW)

Fall 2023

26 / 37

Change of Variables

Distribution Function Technique

Exercise

Theorem [Casella & Berger (2002)]

Let X be a continuous random variable with a *strictly monotone* cumulative distribution function $F_X(x)$. Then, the random variable Y defined as

$$Y:=F_X(X)$$

has a uniform distribution.

Exercise

Consider $f_X(x) = 3x^2$ in the previous example. Show that $Y := F_X(X)$ attains a uniform distribution.

◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ◆ □ ● ○ ○ ○

27 / 37

Fall 2023

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

Change of Variables

Distribution Function Technique

Remark

The first approach relies on the following facts:

- We can transform the cdf of Y into an expression that is a cdf of X.
- We can differentiate the cdf to obtain the pdf.

< □ > < □ > < □ > < □ > < □ > < □ >

Change of Variables

Change of Variables

Outline

Gaussian Distribution

- Marginals and Conditionals of Gaussians
- Sums and Linear Transformations

Change of Variables

2

- Distribution Function Technique
- Change of Variables

< □ > < □ > < □ > < □ > < □ > < □ >

Change of Variables

Change of Variables

What We have Learnt From the Calculus Course

$$\int f(g(x))g'(x)dx = \int f(u)du$$
, where $u = g(x)$.

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

Fall 2023

イロト イボト イヨト イヨト

э

Change of Variables

Change of Variables

What We have Learnt From the Calculus Course

$$\int f(g(x))g'(x)dx = \int f(u)du$$
, where $u = g(x)$.

• Intuitively, considering $du \approx \Delta u = g'(x)\Delta x$ as the "small changes".

イロト 不得 ト イヨト イヨト

э

Change of Variables

Change of Variables

The Roadmap (1/2)

- Consider a univariate random variable X and an invertible function U such that Y := U(X).
- Assume that X has states $x \in [a, b]$.
- By the definition of a cdf, we have

Change of Variables

Change of Variables

The Roadmap (1/2)

- Consider a univariate random variable X and an invertible function U such that Y := U(X).
- Assume that X has states $x \in [a, b]$.
- By the definition of a cdf, we have

$$F_Y(y) = \Pr[Y \leq y]$$

Joseph C. C. Lin (CSIE, TKU, TW)

Change of Variables

Change of Variables

The Roadmap (1/2)

- Consider a univariate random variable X and an invertible function U such that Y := U(X).
- Assume that X has states $x \in [a, b]$.
- By the definition of a cdf, we have

$$F_Y(y) = \Pr[Y \le y] = \Pr[U(X) \le y]$$

ML Math - Probability & Distributions

Fall 2023

< 177 ▶

Change of Variables

Change of Variables

The Roadmap (1/2)

- Consider a univariate random variable X and an invertible function U such that Y := U(X).
- Assume that X has states $x \in [a, b]$.
- By the definition of a cdf, we have

$$F_Y(y) = \Pr[Y \le y] = \Pr[U(X) \le y]$$

If U is strictly increasing, then so is its inverse U^{-1} .

$$\Pr[U(X) \le y] = \Pr[U^{-1}(U(X)) \le U^{-1}(y)]$$

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probab

ML Math - Probability & Distributions

Fall 2023

31 / 37

Change of Variables

Change of Variables

The Roadmap (1/2)

- Consider a univariate random variable X and an invertible function U such that Y := U(X).
- Assume that X has states $x \in [a, b]$.
- By the definition of a cdf, we have

$$F_Y(y) = \Pr[Y \le y] = \Pr[U(X) \le y]$$

If U is strictly increasing, then so is its inverse U^{-1} .

$$\Pr[U(X) \le y] = \Pr[U^{-1}(U(X)) \le U^{-1}(y)] = \Pr[X \le U^{-1}(y)].$$

Then,
$$F_Y(y) = \Pr[X \le U^{-1}(y)] = \int_a^{U^{-1}(y)} f_X(x) dx$$

Joseph C. C. Lin (CSIE, TKU, TW)

Fall 2023

< 17 ▶

31 / 37

Change of Variables

Change of Variables

The Roadmap (2/2)

• To obtain the pdf, we differentiate $F_Y(y)$ w.r.t. y:

$$f_{Y}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{Y}(y) = \frac{\mathrm{d}}{\mathrm{d}y} \int_{a}^{U^{-1}(y)} f_{X}(x) \mathrm{d}x.$$

Change of Variables

Change of Variables

The Roadmap (2/2)

• To obtain the pdf, we differentiate $F_Y(y)$ w.r.t. y:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} \int_a^{U^{-1}(y)} f_X(x) \mathrm{d}x.$$

- The integral on the right-hand side is w.r.t. x, but we need an integral w.r.t. y (∵ we are differentiating w.r.t. y...)
- Change-of-variable comes to the rescue!

Change of Variables

Change of Variables

The Roadmap (2/2)

• To obtain the pdf, we differentiate $F_Y(y)$ w.r.t. y:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} \int_a^{U^{-1}(y)} f_X(x) \mathrm{d}x.$$

- The integral on the right-hand side is w.r.t. x, but we need an integral w.r.t. y (:: we are differentiating w.r.t. y...)
- Change-of-variable comes to the rescue!

•
$$\int f_X(U^{-1}(y))U^{-1'}(y)dy = \int f_X(x)dx$$
, where $x = U^{-1}(y)$.

Change of Variables

Change of Variables

The Roadmap (2/2)

• To obtain the pdf, we differentiate $F_Y(y)$ w.r.t. y:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} \int_a^{U^{-1}(y)} f_X(x) \mathrm{d}x.$$

• The integral on the right-hand side is w.r.t. x, but we need an integral w.r.t. y (: we are differentiating w.r.t. y...)

• Change-of-variable comes to the rescue!

•
$$\int f_X(U^{-1}(y))U^{-1'}(y)dy = \int f_X(x)dx$$
, where $x = U^{-1}(y)$.

$$f_{Y}(y) = \frac{d}{dy} \int_{a}^{U^{-1}(y)} f_{X}(U^{-1}(y)) U^{-1'}(y) dy$$

= $f_{X}(U^{-1}(y)) \cdot \left(\frac{d}{dy} U^{-1}(y)\right).$

Thus,

Change of Variables

Change of Variables

Remark

For decreasing functions,

$$f_Y(y) = -f_X(U^{-1}(y)) \cdot \left(\frac{\mathrm{d}}{\mathrm{d}y}U^{-1}(y)\right)$$

Joseph C. C. Lin (CSIE, TKU, TW) ML M

ML Math - Probability & Distributions

Fall 2023

<ロト <回ト < 回ト < 回ト < 回ト -

33 / 37

æ

ML Math - Probability & Distributions Change of Variables

Change of Variables

Remark

For decreasing functions,

$$f_Y(y) = -f_X(U^{-1}(y)) \cdot \left(\frac{\mathrm{d}}{\mathrm{d}y}U^{-1}(y)\right).$$

So for both increasing and decreasing U,

$$f_Y(y) = f_X(U^{-1}(y)) \cdot \left| \frac{\mathrm{d}}{\mathrm{d}y} U^{-1}(y) \right|.$$

Joseph C. C. Lin (CSIE, TKU, TW) ML Ma

ML Math - Probability & Distributions

Fall 2023

イロト イボト イヨト イヨト

33 / 37

э

ML Math - Probability & Distributions Change of Variables

Change of Variables

Remark

For decreasing functions,

$$f_Y(y) = -f_X(U^{-1}(y)) \cdot \left(\frac{\mathrm{d}}{\mathrm{d}y}U^{-1}(y)\right).$$

So for both increasing and decreasing U,

$$f_Y(y) = f_X(U^{-1}(y)) \cdot \left| \frac{\mathrm{d}}{\mathrm{d}y} U^{-1}(y) \right|.$$

• The term $\left|\frac{d}{dy}U^{-1}(y)\right|$ measures how much a unit volume changes when applying U.

・ロト ・ 同ト ・ ヨト ・ ヨト

Change of Variables

Change of Variables

The Main Theorem

Theorem [Billingsley (1995)]

Let $f_X(\mathbf{x})$ be the pdf of the multivariate continuous random variable X. If the vector-valued function $\mathbf{y} = U(\mathbf{x})$ is differentiable and invertible for all values within the domain of \mathbf{x} , then for corresponding values of \mathbf{y} , the pdf of Y = U(X) is given by

$$f(\mathbf{y}) = f_{\mathbf{x}}(U^{-1}(\mathbf{y})) \cdot \left| \det \left(\frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right) \right|.$$

Fall 2023

34 / 37

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

Change of Variables

Change of Variables

Example

Consider a bivariate random variable X with states $\mathbf{x} = \begin{vmatrix} x_1 \\ x_2 \end{vmatrix}$ and pdf

$$f\left(\left[\begin{array}{c}x_1\\x_2\end{array}\right]\right) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left[\begin{array}{c}x_1\\x_2\end{array}\right]^{\top}\left[\begin{array}{c}x_1\\x_2\end{array}\right]\right)$$

Then, consider a matrix $\textbf{\textit{A}} \in \mathbb{R}^{2 \times 2}$ defined as

$$\mathbf{A} = \left[\begin{array}{cc} a & b \\ c & d \end{array} \right]$$

Goal: Find the pdf of the random variable Y with states $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Joseph C. C. Lin (CSIE, TKU, TW)

ML Math - Probability & Distributions

Fall 2023

イロト 不得 トイヨト イヨト 二日

35 / 37

Change of Variables

Change of Variables

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

▲□▶ ▲圖▶ ▲厘▶ ▲厘▶

2

Change of Variables

Change of Variables

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

2

Change of Variables

Change of Variables

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

3

イロト イヨト イヨト イヨト

Change of Variables

Change of Variables

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

• The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^{\top}(\mathbf{A}^{-1})^{\top}\mathbf{A}^{-1}\mathbf{y}\right)$$

ML Math - Probability & Distributions

イロン イ理 とく ヨン イ ヨン

Fall 2023

Joseph C. C. Lin (CSIE, TKU, TW)

Change of Variables

Change of Variables

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

• The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^{\top}(\mathbf{A}^{-1})^{\top}\mathbf{A}^{-1}\mathbf{y}\right)$$

•
$$\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y} =$$

Joseph C. C. Lin (CSIE, TKU, TW)

ML Math - Probability & Distributions

▶ ≣ つへで 36/37

イロン イ理 とく ヨン イ ヨン

Fall 2023

Change of Variables

Change of Variables

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

• The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^{\top}(\mathbf{A}^{-1})^{\top}\mathbf{A}^{-1}\mathbf{y}\right)$$

•
$$\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y} = \mathbf{A}^{-1}$$
. So, det $\left(\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y}\right) = \det(\mathbf{A}^{-1}) =$

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Pr

ML Math - Probability & Distributions

Fall 2023

イロン イ理 とく ヨン イ ヨン

36 / 37

æ

Change of Variables

Change of Variables

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

• The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^{\top}(\mathbf{A}^{-1})^{\top}\mathbf{A}^{-1}\mathbf{y}\right)$$

•
$$\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y} = \mathbf{A}^{-1}$$
. So, det $\left(\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y}\right) = \det(\mathbf{A}^{-1}) = \frac{1}{ad - bc}$

Joseph C. C. Lin (CSIE, TKU, TW) ML Ma

ML Math - Probability & Distributions

Fall 2023

イロン イ理 とく ヨン イ ヨン

36 / 37

æ

Change of Variables

Change of Variables

•
$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

• The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^{\top}(\mathbf{A}^{-1})^{\top}\mathbf{A}^{-1}\mathbf{y}\right)$$

•
$$\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y} = \mathbf{A}^{-1}$$
. So, det $\left(\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y}\right) = \det(\mathbf{A}^{-1}) = \frac{1}{ad - bc}$
• Thus, $f(\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^{\top}(\mathbf{A}^{-1})^{\top}\mathbf{A}^{-1}\mathbf{y}\right) \cdot \left|\frac{1}{ad - bc}\right|$.

Joseph C. C. Lin (CSIE, TKU, TW)

イロン イ理 とく ヨン イ ヨン

3

Discussions

Joseph C. C. Lin (CSIE, TKU, TW) ML Math - Probability & Distributions

Fall 2023

A D N A B N A B N A B N

э